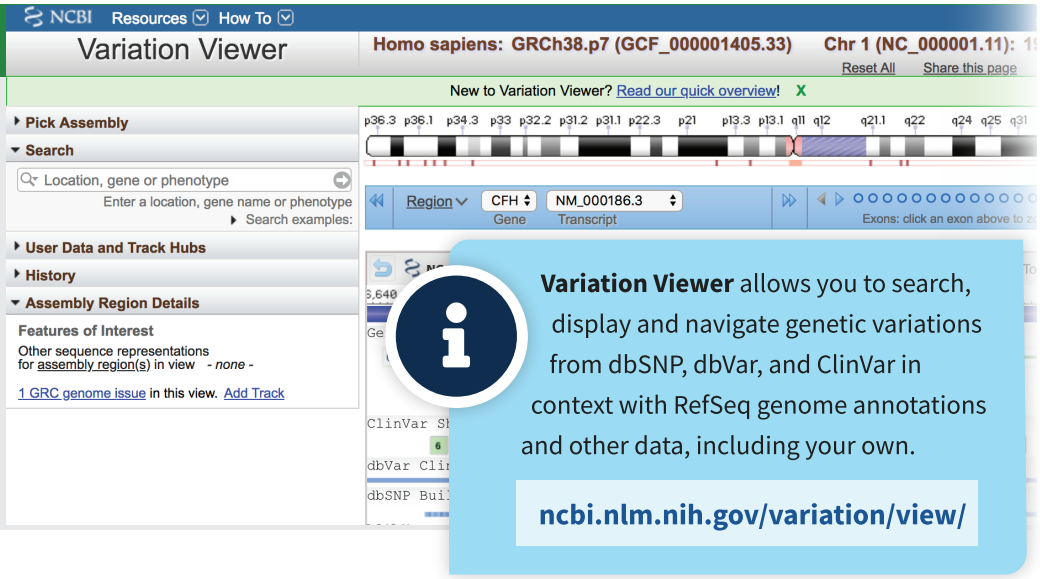
Interactive Visualization of NCBI Annotations for Variant Interpretations

ncbi.nlm.nih.gov/variation/view/

Case Study

A physician has ordered a gene sequencing genetic test for a patient exhibiting immune disruption (suspected *ADA* Deficiency). She has sent you a VCF (with coordinates based on GRCh37) containing three variants and has asked you to evaluate them and prioritize them for further experimental analysis.

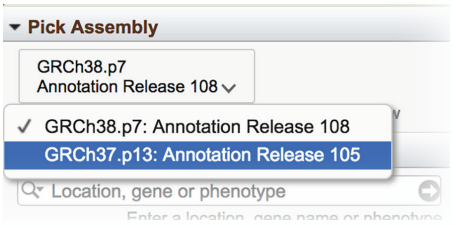


Setting up your data in the NCBI Variation Viewer

Step 1 Go to NCBI Variation Viewer

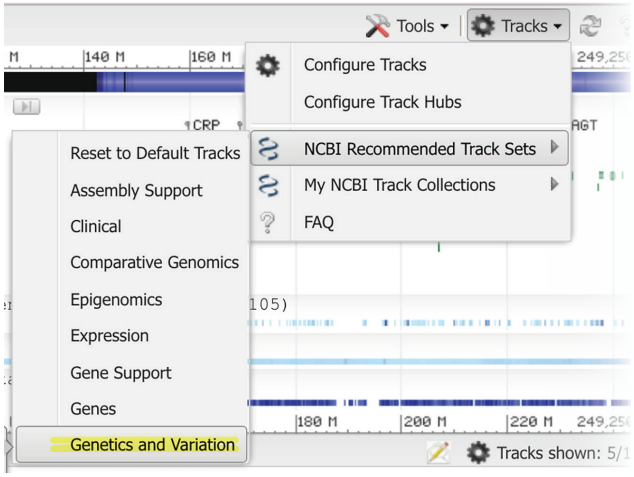
Start in GRCh37 Variation Viewer (ncbi.nlm.nih.gov/variation/view/).

Since many variant analyses are initially done in GRCh37, use the Pick Assembly widget at top left and select to display GRCh37.p13 – Annotation Release 105.



Step 2 Display the *Genetics and Variation* Track Set

Go to *Tracks* on the right of the toolbar, select *NCBI Recommended Track Sets*, and then select *Genetics and Variation*.



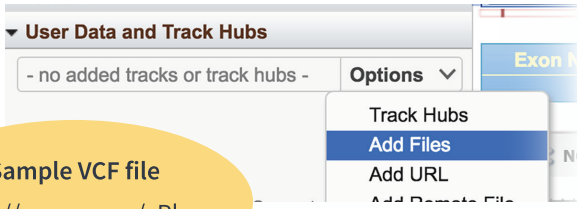
NCBI provides Track Sets (pre-selected groups of high value tracks) to facilitate common analyses. The *Genetics and Variation* track set contains a specified subset of gene and variant-related information. There are many more tracks available in the *Configure Tracks* window.

YouTube **Video of this tutorial at** bit.ly/VariantInterpretation

NIH U.S. National Library of Medicine
National Center for Biotechnology Information

Step 3 — Load VCF file to display the variant calls

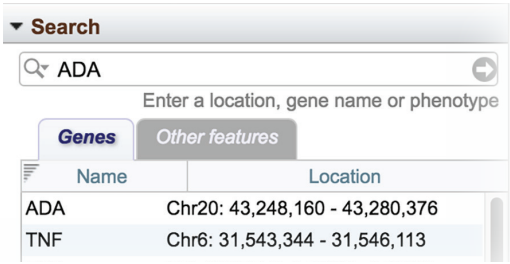
Expand the *User Data and Track Hubs* widget panel on left, from the *Options pull-down* menu, click *Add Files* and select your file. You can name your own track, if desired (e.g. Patient Variants). User tracks are highlighted in green.



Sample VCF file
https://go.usa.gov/xPkya

Step 4 — Visualize the *ADA* gene

Orient yourself to the view shown. Search for *ADA* using the *Search* widget on your left. The display will zoom in the genomic region and show the *ADA's* gene structure.

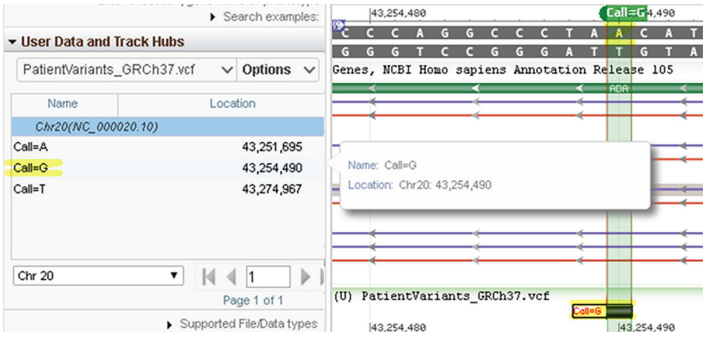


Validate the variant sequence by comparison with the reference genome assembly

Step 5 — Zoom in to a particular variant-of-interest for further analysis

In the *User Data and Track Hubs* widget, click on *Call=G* to navigate to location of this variant.

Note that this "G" was called a variant based on GRCh37 reference assembly where you can see the corresponding reference "A" base.



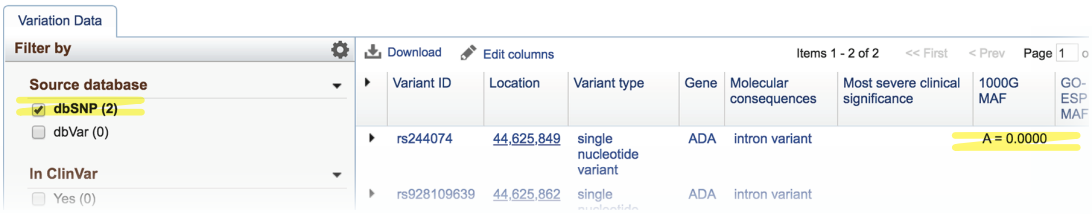
However, with concerns about relying on an older reference assembly, you decide to check the variant against the up-to-date GRCh38.p7 assembly.

Step 6 — Repeat the previous steps for GRCh38.

In the *Pick Assembly* widget, change the selection to GRCh38.p7 – Annotation Release 108. Load the VCF file that has been updated based on GRCh38 coordinates in the *User Data and Track Hubs* widget. Then, click on *Call=G* again to navigate to the location of the variant and compare your "G" call to what is shown in the GRCh38 reference. In this case, it is also shown a a "G".

Sample VCF file
https://go.usa.gov/xPkyD

To learn more about the frequency of the "A" call at this position, scroll down to the Variant Table shown below the graphical display. Filter the table by clicking *dbSNP* to show two known variants in this region. In the row corresponding to the Call=G variant (rs244704), a minor allele frequency (MAF) determined by the 1000 Genomes (1000G) Project based on the GRCh37 assembly is listed as 0.000. This is indicative of a possible GRCh37 assembly error as this indicates that "A" is never found in 1000 Genomes dataset, which is unlikely.



You conclude that this "G" is not a "variant" after all, so there is no need to pursue further & you decide to continue analysis using the current GRCh38 assembly from this point forward.

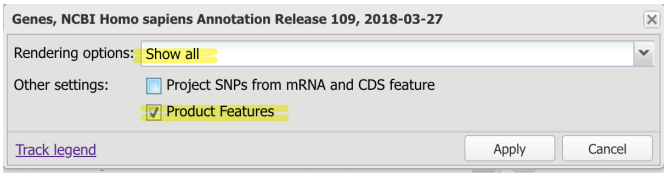
Identify the position of the variant and assess for impact in well-annotated protein coding regions

Step 7 Zoom in to a particular variant-of-interest for further analysis

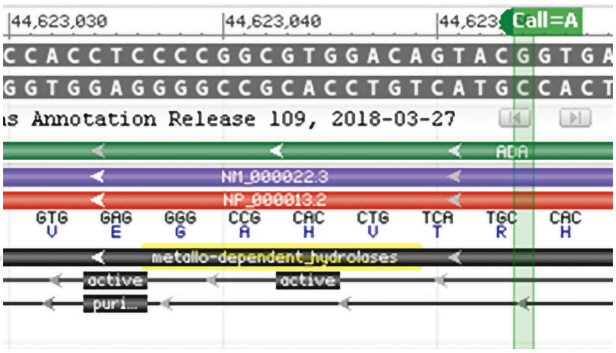
In the *User Data and Track Hubs* widget, click on *Call=A* to navigate to the location of this variant. Note that the reference genome contains a "G" in this location, so the Call=A appears to be correctly called as a variant.

Step 8 Change display of Gene track to expand and show features

To examine the gene context of this particular variant, expand the gene track to show all gene products (mRNA transcripts and proteins) and show annotated features. Right-click over track title to open track tool tip. Click left-most cog icon and select Rendering Option > Show All and Project Features.



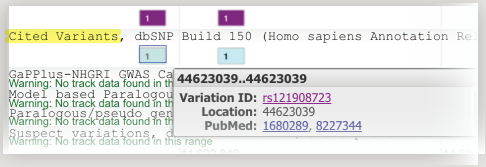
You note that this variant SNP is located in a metallo-dependent_hydrolase region and very close to key active site residues within the annotated conserved domain.



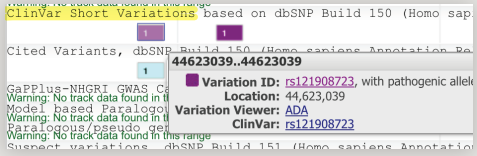
Other helpful tracks

Hover over these tracks for more information.

Cited Variants - displays SNP-listed variants with a link to related PubMed entries.



ClinVar Short Variants - displays information and links to more data in the ClinVar database.



You conclude that this "A" is a good variant for further experimentation (a high priority) due to its location near an active site of an annotated protein domain.

Identify the position of the variant and assess for impact in newly-annotated non-coding regions.

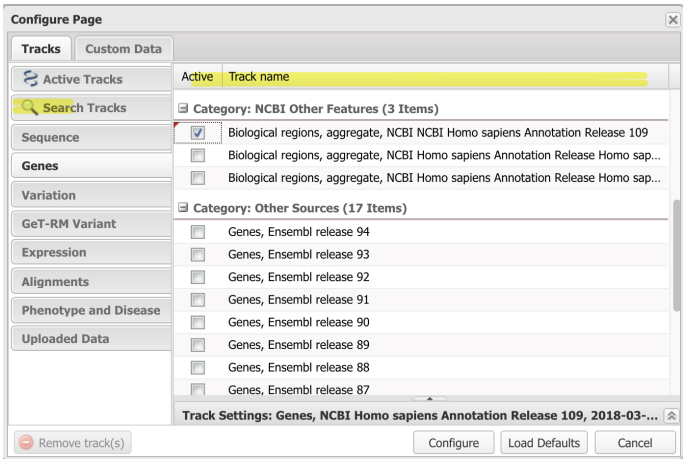
Step 9 Zoom in to another particular variant-of-interst for further analysis

In the *User Data and Track Hubs* widget, click on *Call=C* to navigate to the location of the variant. Note that the reference genome contains a "A" in this location, so the Call=C appears to be correctly called as a variant.

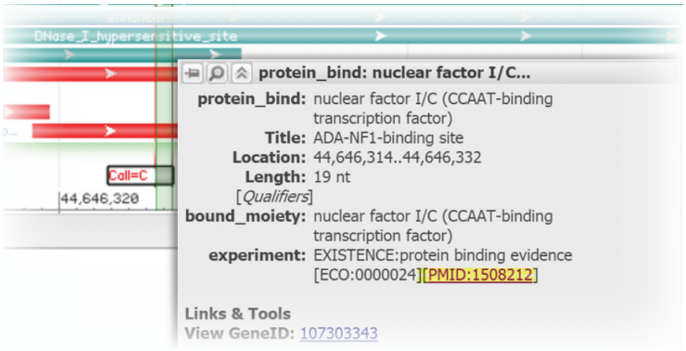
In this gene-context view you note that this variant is located in an intronic region and there appears to be no variant-related annotations. However, before giving up on learning more about this variant, you decide to add NCBI's newest RefSeq annotation track to the display: *Biological regions.*

Step 10 Add *Biological regions* track to the display

On the toolbar, go to *Tracks* > *Configure Tracks*. In the track categories on the left, click *Genes*. To add the *Biological Regions* track, scroll down to *Category: NCBI Other Features (3 Items)*, and check *Biological regions, aggregate, NCBI NCBI Homo sapiens Annotation Release 109*, then click the *Configure* button on the bottom of the window.



You note that this new track contains information about a Thymic "locus_control_region" and "enhancer", as well as transcription factor binding sites. Of particular interest is "protein_bind: nuclear factor I…." with a PubMed citation where you can learn more by following the PMID link.



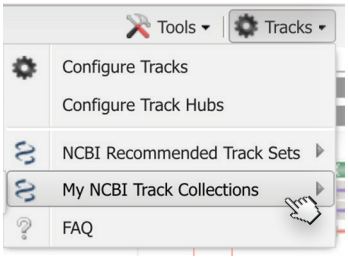
Conclusion

You conclude that this "T" variant could impact gene expression by affecting the NF1 transcription factor binding site and is worthy of further experimental analysis.

Other Variation Viewer features

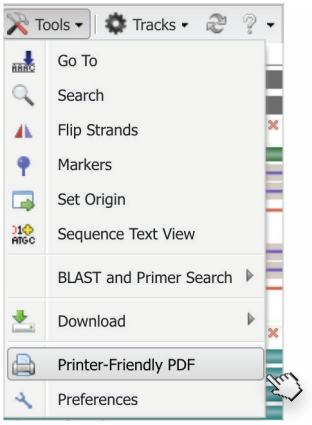
Save your custom configured display

In the toolbar, go to Tracks > My NCBI Track Collections. It is a quick way to load your custom display configurations! This requires an NCBI account.



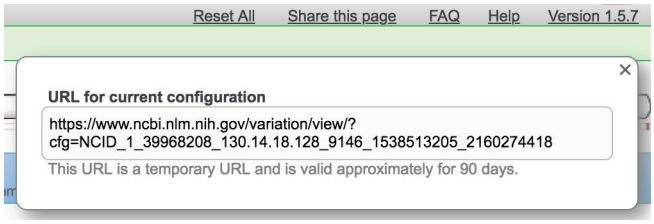
Create a publication quality image

Again, in the toolbar, under *Tools* menu the Printer-Friendly PDF option gives you a good quality image that can be used in publications.



Share with a collaborator

At the top right corner, click on *Share this page* to get a link with your current configuration you can send to others.



Need help? Email us at info@ncbi.nlm.nih.gov

U.S. National Library of Medicine
National Center for Biotechnology Information

Video of this tutorial at bit.ly/VariantInterpretation